

# Entropy Distance: New Quantum Phenomena

Andreas Knauf<sup>1</sup> and Stephan Weis<sup>2</sup>

Department Mathematik, Friedrich-Alexander-Universität Erlangen-Nürnberg, Bismarckstraße  
1 $\frac{1}{2}$ , D-91054 Erlangen, Germany.

August 27, 2010

**Abstract** – The relative entropy distance of a state from an exponential family is important in information theory and statistics. The class of exponential families is parametrized by a Grassmannian. A minimal example in the algebra of complex 3x3 matrices shows that the mean value set of an exponential family has typically non-exposed faces. Where non-exposed faces are born in the Grassmannian, families have a discontinuous entropy distance.

These two phenomena are related to three distinct closures, which all coincide in the probabilistic case of the algebra  $\mathbb{C}^N$ . A necessary condition for a local maximizer of the entropy distance is calculated in a finite-dimensional complex matrix algebra.

*Index Terms* – exponential family, relative entropy.

*AMS Subject Classification:* 52A10, 81P45, 94A17.

## 1 Introduction

### 1.1 Overview

In this article we consider special submanifolds  $\mathcal{E}$  of state space  $\overline{\mathcal{S}}(\mathcal{A})$  for algebras  $\mathcal{A}$  of observables.  $\mathcal{A}$  itself is called *classical* or *probabilistic* if it is abelian and *quantum* otherwise. These *exponential families* are images of the real affine subspaces of the algebra of observables under the trace-normalized exponential map. They are important in information theory, in statistics, in statistical physics and in neural networks, to name just a few fields. See Amari and Nagaoka [AN00] for an overview.

It is well known that classical and quantum state spaces have quite different geometries. Whereas in the classical case we have a simplex (and thus every state is uniquely decomposed into pure states), in the quantum case such a decomposition is highly non-unique (think of the Bloch ball  $\overline{\mathcal{S}}(\mathcal{A})$  for  $\mathcal{A} = \text{Mat}(2, \mathbb{C})$ ).

Still from the point of view of convex geometry there is one common property of all these state spaces: all of their faces are exposed faces<sup>3</sup>, that is, they can be described as the intersection of state space with a half space.

In the probabilistic setting of  $\mathcal{A} = \mathbb{C}^N$  measurement of observables  $f_1, \dots, f_n$  leads to an orthogonal projection

$$\overline{\mathcal{S}}(\mathcal{A}) \longrightarrow \mathbb{R}^n, \quad p \longmapsto (\mathbb{E}_p(f_1), \dots, \mathbb{E}_p(f_n))$$

of state space, based on expectation  $\mathbb{E}_p$ . The image, called *mean value set* or *convex support* is no longer a simplex but still a polygon. So faces of a classical mean value set are exposed faces, too. The same applies to all exponential families and their natural projections, see Figure 1.

We exhibit here two main differences between the abelian and the non-abelian case of an exponential family, at least in our minimal example.

1. First, in the quantum case it is typical in a Grassmann manifold parametrizing the families that mean value sets of an exponential family have non-exposed faces.
2. Second, the entropy distance from an exponential family defined in (7) can be discontinuous in exceptional cases. The boundary in the Grassmann manifold dividing non-exposed faces from exposed faces seems pivotal for discontinuity.

---

<sup>1</sup>knauf@mi.uni-erlangen.de

<sup>2</sup>weis@mi.uni-erlangen.de

<sup>3</sup>*non-exposed faces* are found, e.g., on the circumference of a stadium, at the four points where a half-circle meets a segment. See Section 2.2 for precise definitions.

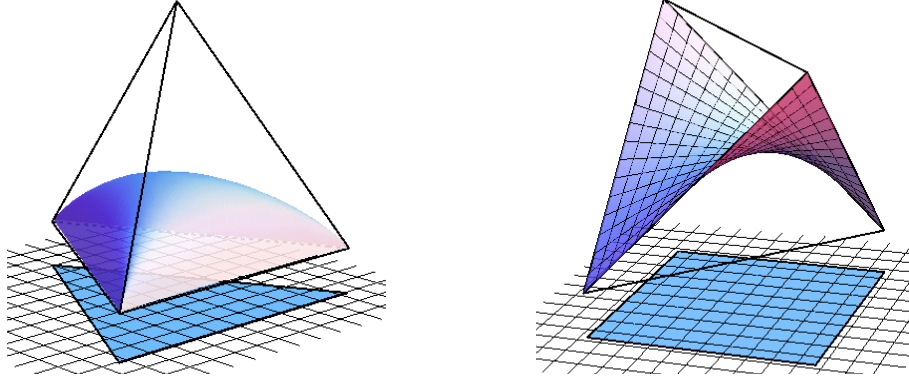


Figure 1: Mean value sets for two probabilistic exponential families. Left: triangle; right: square.

We build on the work of Wichmann, who studied ensembles of maximum chaos in  $\mathcal{A} = \text{Mat}(N, \mathbb{C})$  in 1963. We borrow from this source the mean value chart, which we will introduce in Section 1.2. However, Wichmann ignored the existence of non-exposed faces at a quantum mean value set: Theorem I (e) in [Wi63] concerning extreme points is wrong, it fails in all cases where non-exposed faces appear.

Non-exposed faces are not a minor issue to the study of entropy distance. First, the simplest possible example of a metamorphosis between a classical (simplicial) and a quantum mean value set convinces that non-exposed faces are typical. This is depicted in Figure 2 and explained in Section 2. Second, the metamorphosis contains one shape that divides mean value sets with non-exposed faces from mean value sets without non-exposed faces. This shape corresponds to the Staffelberg family in Section 3.2 having a discontinuous entropy distance. In this example the convex geometric notion of non-exposed face indicates discontinuity of the entropy distance.

Another reason to study the above-mentioned phenomena are their counterparts in the structure of extensions of exponential families. These have been developed in the classical context of a Borel measure on  $\mathbb{R}^d$  by Csiszár and Matúš [CM03, CM05]. For a measure of finite support their results become rather easy and were already investigated by Barndorff-Nielsen [Ba78] and Čencov [Ce82]. The probabilistic case of finite support corresponds to the algebra  $\mathcal{A} = \mathbb{C}^N$  where the following extensions all coincide.

Three extensions of  $\mathcal{E}$  have been worked out by Weis [We09] in the context of a  $*$ -subalgebra  $\mathcal{A}$  of  $\text{Mat}(N, \mathbb{C})$ . The *reverse information closure* or *rl-closure*

$$\text{cl}_{\text{rl}}(\mathcal{E}) := \{\rho \in \overline{\mathcal{S}}(\mathcal{A}) \mid d_{\mathcal{E}}(\rho) = 0\}$$

consists of states that approximate  $\mathcal{E}$  in relative entropy distance (7) and the *norm closure*<sup>4</sup>  $\overline{\mathcal{E}}$  is taken in the Hilbert-Schmidt norm topology of  $\mathcal{A}$ . These closures were used by Csiszár and Matúš [CM05] but the inclusion  $\text{cl}_{\text{rl}}(\mathcal{E}) \subset \overline{\mathcal{E}}$  holds also in the quantum case. It arises from the Pinsker-Csiszár inequality, see e.g. Petz [Pe08]. To our best knowledge the *geodesic closure*<sup>5</sup>

$$\text{cl}_{\text{geo}}(\mathcal{E}) := \mathcal{E} \cup \{\text{limit points of e-geodesics in } \mathcal{E}\} \quad (1)$$

is new in the literature. In Section 3 we prove the inclusions

$$\text{cl}_{\text{geo}}(\mathcal{E}) \subset \text{cl}_{\text{rl}}(\mathcal{E}) \subset \overline{\mathcal{E}}. \quad (2)$$

Equality conditions in Corollary 7.7 and Theorem 7 in [We09] say that  $\text{cl}_{\text{geo}}(\mathcal{E}) = \text{cl}_{\text{rl}}(\mathcal{E})$  if and only if the mean value set has no non-exposed faces and  $\text{cl}_{\text{rl}}(\mathcal{E}) = \overline{\mathcal{E}}$  if and only if  $d_{\mathcal{E}}$  is continuous. Returning to the classical case of  $\mathcal{A} = \mathbb{C}^N$  the equality  $\text{cl}_{\text{geo}}(\mathcal{E}) = \overline{\mathcal{E}}$  holds (see Proposition 8.2 in [We09]) and we arrive at the continuity of  $d_{\mathcal{E}}$  proved by Ay [Ay02].

<sup>4</sup>In probability theory, this is often called *variation closure* and taken in  $L^1$  norm with respect to a dominating measure but in finite dimensions the norm does not matter.

<sup>5</sup>A one-dimensional exponential family is called *e-geodesic*.



Figure 2: A classical—quantum metamorphosis of 2D mean value sets in the algebra  $\text{Mat}(2, \mathbb{C}) \oplus \mathbb{C}$ . On the left we have the triangle  $\overline{\mathcal{S}}(\mathbb{C}^3)$  followed by three shapes with non-exposed faces and with a corner of decreasing distinctness. These are followed by three ellipses including on the right a projection of the Bloch ball  $\overline{\mathcal{S}}(\text{Mat}(2, \mathbb{C}))$ . The fifth mean value set divides these with a corner from the ellipses. It belongs to the Staffelfberg family, which has a discontinuous entropy distance.

An example of an exponential family with non-exposed faces but, contrary to the Staffelfberg family, with a continuous entropy distance is provided in Section 3.3 with the swallow family.

Directional derivatives in Section 4 provide a necessary optimality condition. A local maximizer of  $d_{\mathcal{E}}$  proves to be completely determined by its projection onto  $\mathcal{E}$ . This is a quantum analogue of the probabilistic case, where a local maximizer is the truncation of its projection.

## 1.2 The advantage of mean value sets

The entropy distance from an exponential family can be a measure of information, see e.g. Ay and Knauf [AK06] and references therein—an example is multi-information, which measures stochastic dependencies between units of a composite system. We recall how the Pythagorean theorem of relative entropy defines a state of minimal relative entropy on an exponential family. This leads us to the mean value chart and to mean value sets.

For  $N \in \mathbb{N}$  we denote by  $\mathcal{A}$  a  $*$ -subalgebra of  $\text{Mat}(N, \mathbb{C})$  with identity  $\mathbb{1}$ . We write  $a^*$  for the adjoint of  $a \in \mathcal{A}$  and  $\mathcal{A}_{\text{sa}} := \{a \in \mathcal{A} \mid a^* = a\}$  for the real vector space of self-adjoint matrices. The Hilbert-Schmidt inner product of  $a, b \in \mathcal{A}_{\text{sa}}$  is  $\langle a, b \rangle = \text{tr}(ab)$ . A matrix  $p \in \mathcal{A}$  is an *orthogonal projector* if  $p^2 = p^* = p$ . A matrix  $a \in \mathcal{A}$  is *positive*, written as  $a \geq 0$ , if  $a^* = a$  and  $a$  has no negative eigenvalue. The *state space* of  $\mathcal{A}$  is  $\overline{\mathcal{S}}(\mathcal{A}) := \{\rho \in \mathcal{A}_{\text{sa}} \mid \rho \geq 0, \text{tr}(\rho) = 1\}$ , its elements are *states*.

The space of invertible states is  $\mathcal{S}(\mathcal{A}) := \{\rho \in \overline{\mathcal{S}}(\mathcal{A}) \mid \rho^{-1} \text{ exists}\}$ . If  $\mathcal{A} \neq \{0\}$  then the *trace-normalized exponential* is the real analytic mapping

$$\exp_1 : \mathcal{A}_{\text{sa}} \longrightarrow \mathcal{S}(\mathcal{A}), \quad a \longmapsto \frac{e^a}{\text{tr}(e^a)}.$$

This is a diffeomorphism onto  $\mathcal{S}(\mathcal{A})$  while restricted to traceless matrices. The real analytic inverse  $\ln_0 : \rho \mapsto \ln(\rho) - \mathbb{1} \text{tr}(\ln(\rho))/\text{tr}(\mathbb{1})$  is the *canonical chart* of  $\mathcal{S}(\mathcal{A})$ . The image of a non-empty affine subspace of  $\mathcal{A}_{\text{sa}}$  under  $\exp_1$  is an *exponential family* in  $\mathcal{A}$ . For an exponential family  $\mathcal{E}$  we call  $\ln_0|_{\mathcal{E}}$  the *canonical chart* of  $\mathcal{E}$  and the affine space  $\Theta := \ln_0(\mathcal{E})$  the *canonical parameter space* of  $\mathcal{E}$ . The translation vector space  $V$  of  $\Theta$  is the *canonical tangent space* of  $\mathcal{E}$ . Sometimes we consider *linear exponential families* with  $\Theta = V$  to use preparatory work done by Wichmann [Wi63].

The *relative entropy* between states  $\rho, \sigma \in \overline{\mathcal{S}}(\mathcal{A})$  is  $S(\rho, \sigma) := +\infty$  unless  $\sigma$  has the larger image  $\text{Im}(\sigma) \supset \text{Im}(\rho)$  and then (using the natural logarithm)

$$S(\rho, \sigma) := \text{tr} \rho (\ln(\rho) - \ln(\sigma)). \quad (3)$$

The distance-like properties of  $S(\rho, \sigma) \geq 0$  and of  $S(\rho, \sigma) = 0 \iff \rho = \sigma$  hold, cf. Wehrl [We78].

The relative entropy is very well compatible with exponential families. If  $\rho, \sigma$  and  $\tau$  are states in  $\mathcal{A}$  with  $\sigma$  and  $\tau$  invertible, and if  $\rho - \sigma \perp \ln(\tau) - \ln(\sigma)$ , then we have

$$S(\rho, \sigma) + S(\sigma, \tau) = S(\rho, \tau). \quad (4)$$

This is the *Pythagorean theorem of relative entropy*<sup>6</sup>. It is easy to see that the Pythagorean theorem (4) holds if  $\sigma$  and  $\tau$  belong to an exponential family  $\mathcal{E}$  in  $\mathcal{A}$  and if  $\rho \in \overline{\mathcal{S}}(\mathcal{A})$  satisfies  $\rho - \sigma \perp V$ . Therefore we have a projection for states  $\rho \in \mathcal{E} + V^\perp$  defined by intersection  $\pi_{\mathcal{E}}(\rho) = (\rho + V^\perp) \cap \mathcal{E}$  understood as the element of a one-element set. This projection satisfies

$$\inf_{\sigma \in \mathcal{E}} S(\rho, \sigma) = \min_{\sigma \in \mathcal{E}} S(\rho, \sigma) = S(\rho, \pi_{\mathcal{E}}(\rho)) \quad (5)$$

and it extends to the cylinder  $\mathcal{E} + V^\perp$  of self-adjoint matrices

$$\pi_{\mathcal{E}} : \mathcal{E} + V^\perp \longrightarrow \mathcal{E}, \quad a \longmapsto (a + V^\perp) \cap \mathcal{E}. \quad (6)$$

The infimum of relative entropy distance (3) between a fixed state and members of the exponential family  $\mathcal{E}$  is called the *entropy distance* from  $\mathcal{E}$

$$d_{\mathcal{E}} : \overline{\mathcal{S}}(\mathcal{A}) \longrightarrow \mathbb{R}, \quad \rho \longmapsto \inf_{\sigma \in \mathcal{E}} S(\rho, \sigma). \quad (7)$$

This is a non-negative and bounded real function on  $\overline{\mathcal{S}}(\mathcal{A})$ .

The question arises how large the orthogonal projection of  $\mathcal{E}$  to  $V$  is. Wichmann has solved this problem in the context of maximum entropy ensembles under linear constraints in Theorem II of [Wi63]. We cite the relevant part of this theorem starting with linear independent observables  $\mathbb{1}, X_1, \dots, X_p \in \mathcal{A}_{\text{sa}}$ . The set of *simultaneous mean values* consists of the  $p$ -tuples of real numbers  $W := \{(\langle X_1, \rho \rangle, \dots, \langle X_p, \rho \rangle) \mid \rho \in \overline{\mathcal{S}}(\mathcal{A})\}$ . There is a mapping

$$\omega : \mathbb{R}^p \longrightarrow \mathcal{S}(\mathcal{A}), \quad y \longmapsto \exp_1 \left( \sum_{i=1}^p y_i X_i \right)$$

and a second mapping

$$x : \mathcal{A}_{\text{sa}} \longrightarrow \mathbb{R}^p, \quad \omega \longmapsto (\langle X_1, \omega \rangle, \dots, \langle X_p, \omega \rangle),$$

such that  $x \circ \omega : \mathbb{R}^p \rightarrow \mathbb{R}^p$ ,  $y \mapsto x(\omega(y))$  is real analytic,  $x \circ \omega$  is injective and onto the interior of the set of simultaneous mean values  $W$ .

We use Wichmanns results in a coordinate free version. They apply to a linear exponential family  $\mathcal{E}$  with canonical tangent space  $V = \Theta = \ln_0(\mathcal{E})$  as follows. The orthogonal projection  $\pi_V : \mathcal{A}_{\text{sa}} \rightarrow V$  onto  $V$  with respect to the Hilbert-Schmidt inner product defines the *mean value set*

$$\text{mv}(V) := \pi_V(\overline{\mathcal{S}}(\mathcal{A})). \quad (8)$$

The *mean value chart* for  $\mathcal{E}$  is the bijection

$$\pi_V|_{\mathcal{E}} : \mathcal{E} \longrightarrow \text{Int}(\text{mv}(V)) \quad (9)$$

onto the interior of the mean value set (open in the topology of  $V$ ). The chart change  $\pi_V|_{\mathcal{E}} \circ \exp_1|_V$  from the canonical chart to the mean value chart is real analytic.

Moreover, Wichmann proved in Lemma VI, [Wi63], that the mapping  $x \circ \omega$  has a positive definite Jacobian. This translates to the positive definite Jacobian of  $\pi_V|_{\mathcal{E}} \circ \exp_1|_V$  and then the *real analytic inverse function theorem* proved by Krantz and Parks [KP02] implies that the opposite chart change  $\ln_0|_{\mathcal{E}} \circ (\pi_V|_{\mathcal{E}})^{-1}$  is real analytic. Composition with  $\exp_1$  from the left and  $\pi_V$  from the right yields the real analytic projection  $\pi_{\mathcal{E}} = (\pi_V|_{\mathcal{E}})^{-1} \circ \pi_V|_{\mathcal{E}+V^\perp}$  defined in (6).

## 2 A classical—quantum metamorphosis

Non-exposed faces appear only in the projection of a state space. So, in a sense, the convex geometry of a mean value set is more complicated than the geometry of a state space. We discuss an example where non-exposed faces are typical. This is made precise in the final discussion of the section. The example is minimal in two respects:

<sup>6</sup>In more sophisticated terms it states (see e.g. Petz [Pe08]): if the segment  $[\rho, \sigma]$  and the e-geodesic from  $\sigma$  to  $\tau$  meet at  $\sigma$  orthogonally with respect to the BKM metric, then (4) holds.

- The algebra  $\mathcal{A} := \text{Mat}(2, \mathbb{C}) \oplus \mathbb{C}$  is the smallest algebra that allows for a mean value set with non-exposed faces. If an algebra has no subalgebra isomorphic to  $\text{Mat}(2, \mathbb{C})$ , then it is abelian and has a simplex as a state space—the mean value sets are polytopes without non-exposed faces. The algebra  $\text{Mat}(2, \mathbb{C})$  by itself has the 3D Bloch ball as state space, whose projections are a point, a segment or a disk.
- On the other hand, for the projection shape we consider the minimal dimension two that allows for non-exposed faces.

We denote the identity resp. zero in  $\text{Mat}(2, \mathbb{C})$  by  $\mathbb{1}_2$  resp.  $0_2$ . The Pauli  $\sigma$ -matrices are  $\sigma_1 := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ ,  $\sigma_2 := \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$  and  $\sigma_3 := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ . We use the vector of Pauli  $\sigma$ -matrices  $\hat{\sigma} := (\sigma_1, \sigma_2, \sigma_3)$  and for  $b = (b_1, b_2, b_3) \in \mathbb{R}^3$  we write  $b\hat{\sigma} = b_1\sigma_1 + b_2\sigma_2 + b_3\sigma_3$ .

## 2.1 A cone substitute reduces dimension

The state space of  $\mathcal{A} := \text{Mat}(2, \mathbb{C}) \oplus \mathbb{C}$  of our minimal example is a 4D cone with the Bloch ball as its base:<sup>7</sup>

$$\overline{\mathcal{S}}(\mathcal{A}) = \text{conv}(\overline{\mathcal{S}}(\text{Mat}(2, \mathbb{C})) \oplus 0, 0_2 \oplus 1).$$

In order to sharpen our geometric insight and simplify arguments we want to further reduce the dimension. Indeed we can replace it by a 3D cone. The reason is that all 2D linear exponential families are contained in a 3D cone, which may, by an automorphism, be assumed to be in the form (14) below.

The special unitary group  $\text{SU}(2)$  acts in a double cover  $\text{SU}(2) \rightarrow \text{SO}(3)$  on  $\text{span}_{\mathbb{R}}\{\sigma_1, \sigma_2, \sigma_3\}$  by rotation (see e.g. Knörrer [Kn06]). The group  $\text{SO}(3)$  acts transitively on the 2D planes in  $\text{span}_{\mathbb{R}}\{\sigma_1, \sigma_2, \sigma_3\} \oplus 0$ . We choose the reference plane  $W := \text{span}_{\mathbb{R}}\{\sigma_1, \sigma_2\} \oplus 0$ .

Let us write  $z := -\frac{1}{2}\mathbb{1}_2 \oplus 1$ . We denote the space of traceless self-adjoint matrices in  $\mathcal{A}$  by  $X := \text{span}_{\mathbb{R}}\{\sigma_1, \sigma_2, \sigma_3\} \oplus 0 + \mathbb{R}z$ , we denote the Grassmannian manifold of 2D planes in  $X$  by  $\mathbb{G}$ . The group  $\text{SU}(2)$  acts on  $X$  with the identity mapping in the second summand, and it acts on the Grassmannian  $\mathbb{G}$ . It is clear that every 2D plane  $V \in \mathbb{G}$  is included in  $\widetilde{W} + \mathbb{R}z$  for a 2D plane  $\widetilde{W} \subset \text{span}_{\mathbb{R}}\{\sigma_1, \sigma_2, \sigma_3\} \oplus 0$ . By the transitive action of  $\text{SU}(2)$  on these spaces  $\widetilde{W}$ , every orbit of  $\text{SU}(2)$  on  $\mathbb{G}$  meets a 2D plane of

$$U := W + \mathbb{R}z. \quad (10)$$

Therefore, if  $\text{SO}(2)$  acts on  $U$  by rotation of  $W$ , we have the orbit isomorphism

$$\mathbb{G}/\text{SU}(2) \cong \{ \text{2D subspaces of } U \} / \text{SO}(2). \quad (11)$$

A complete orbit invariant is the angle between a 2D plane  $V \in \mathbb{G}$  resp.  $V \subset U$  and the vector  $z$

$$\varphi = \varphi(V) := \angle(V, z). \quad (12)$$

Using (11) we confine attention to 2D subspaces  $V \subset U$  and substitute the 4D cone  $\overline{\mathcal{S}}(\mathcal{A})$  by a 3D cone. We begin with the 2D *base disk*  $K := \overline{\mathcal{S}}(\mathcal{A}) \cap (\frac{1}{2}\mathbb{1}_2 \oplus 0 + W)$  of diameter  $\sqrt{2}$  centered about  $\frac{1}{2}\mathbb{1}_2 \oplus 0$ . We use the vector  $b(\alpha) := (\sin(\alpha), \cos(\alpha), 0)$  and describe the *base circle* of the base disk  $K$  by the pure states

$$\rho(\alpha) := \frac{1}{2}(\mathbb{1}_2 + b(\alpha)\hat{\sigma}) \oplus 0, \quad \alpha \in \mathbb{R}. \quad (13)$$

We build on  $K$  the 3D cone with apex  $0_2 \oplus 1 = \frac{1}{2}\mathbb{1}_2 \oplus 0 + z$

$$C := \text{conv}(K, 0_2 \oplus 1). \quad (14)$$

The cone  $C$  is rotationally symmetric about the line  $l$  through  $\frac{1}{2}\mathbb{1}_2 \oplus 0$  and  $0_2 \oplus 1$ . Every plane in  $\frac{1}{2}\mathbb{1}_2 \oplus 0 + U$  that contains  $l$  meets  $C$  in an equilateral triangle, whose center is the *tracial state*  $\frac{1}{3}\mathbb{1}$ . The triangle is isometrically isomorphic to the state space of  $\mathbb{C}^3$ .

The cone  $C$  is a threefold paradigm of the state space. We have

$$\pi_U(\overline{\mathcal{S}}(\mathcal{A})) = C - \frac{1}{3}\mathbb{1}, \quad (15)$$

---

<sup>7</sup> $\text{conv}$  denotes the convex hull.

$$\overline{\mathcal{S}}(\mathcal{A}) \cap [\tfrac{1}{3}\mathbb{1} + U] = C \quad (16)$$

and

$$C = \overline{\exp_1(U)}. \quad (17)$$

The proofs of these equations are straight forward. In this section only (15) is important. But since  $C$  is the closure of an exponential family, we can study in the next section exponential families inside of  $C$  by taking the intersection (16).

## 2.2 Tangent lines to ellipses produce non-exposed faces

We use two distinct concepts of face to describe the convex geometry of a mean value set. We apply these terms to mean value sets on planes of the Grassmannian  $\mathbb{G}$  discussed in Section 2.1. The second and more general idea about a face will be introduced in two (equivalent) forms.

Let  $M$  be a non-empty compact and convex subset of a Euclidean vector space  $(\mathbb{E}, \langle \cdot, \cdot \rangle)$ . Then for non-zero  $u \in \mathbb{E}$  the *supporting hyperplane* is defined by

$$H(M, u) := \{x \in \mathbb{E} \mid \langle x, u \rangle = \max_{y \in M} \langle y, u \rangle\}.$$

In two dimensions  $\dim(\mathbb{E}) = 2$  this is a line  $l$  that meets  $M$  such that  $M \setminus l$  is convex. A subset  $F$  of  $M$  is an *exposed face* of  $M$  if  $F$  is the intersection of  $M$  with a supporting hyperplane

$$F(M, u) := M \cap H(M, u). \quad (18)$$

$F = \emptyset$  and  $F = M$  are exposed faces by definition.

An important example is the state space of a \*-subalgebra  $\mathcal{A}$  of  $\text{Mat}(N, \mathbb{C})$ . For a self-adjoint matrix  $u \in \mathcal{A}_{\text{sa}}$  we denote by  $\mu_+(u)$  the maximal eigenvalue of  $u$  and by  $p_+(u)$  the spectral projector of  $u$  corresponding to  $\mu_+(u)$ , which we call the *maximal projector* of  $u$ .

**Lemma 2.1.** *If  $u \in \mathcal{A}_{\text{sa}}$  is a non-zero self-adjoint matrix, then the exposed face  $F(\overline{\mathcal{S}}(\mathcal{A}), u)$  consists of the states  $\rho \in \overline{\mathcal{S}}(\mathcal{A})$  such that  $\langle \rho, u \rangle = \mu_+(u)$  or, equivalently,  $\text{Im}(\rho) \subset \text{Im}(p_+(u))$ .*

*Proof:* This follows from the general situation in a C\*-algebra, see e.g. Alfsen and Shultz [AS01], and it is easy to verify this directly for  $\mathcal{A}$ , see e.g. Weis [We09].  $\square$

**Remark 2.2.** The exposed face  $F(\overline{\mathcal{S}}(\mathcal{A}), u)$  in Lemma 2.1 is the state space of the algebra  $p\mathcal{A}p$  for  $p := p_+(u)$  discussed below (22).

Grünbaum [Gr67] calls a subset  $F$  of  $M$  a *poonem*, provided there exist sets  $F_0, \dots, F_n$  such that  $F_0 = F$ ,  $F_n = M$ , and  $F_{i-1}$  is an exposed face of  $F_i$  for  $i = 1, \dots, n$ . A convex subset  $F$  of  $M$  is a *face* of  $M$ , if for all  $x, y \in M$  and all  $0 < \lambda < 1$  the inclusion of  $(1 - \lambda)x + \lambda y \in F$  implies  $x, y \in F$ . A face of  $M$ , which is not an exposed face of  $M$  will be called a *non-exposed face* of  $M$ , see Figure 3.

**Remark 2.3.** For a convex set of finite dimension the concept of poonem is equivalent to the concept of face. This is easy to prove by induction taking an infimum in the complete lattice of exposed faces, see Lemma 5.30 in [We09]. The concept of face is more popular, see e.g. Rockafellar [Ro72], but the concept of poonem is more suitable to study entropy distance, see Remark 3.13.

We return to the Grassmannian  $\mathbb{G}$  discussed in Section 2.1 and consider a 2D subspace  $V$  of the space  $U$  defined in (10) having an angle of  $\varphi = \angle(V, z)$  with the vector  $z$ . By (15) the mean value set is

$$\text{mv}(V) = \pi_V(\overline{\mathcal{S}}(\mathcal{A})) = \pi_V(C)$$

for the 3D cone  $C = \text{conv}(K, 0_2 \oplus 1)$ . Here  $K$  is a disk with symmetry axis in the direction of  $z$ . The mean value set is the convex hull of  $k := \pi_V(K)$  and of the projection  $x := \pi_V(0_2 \oplus 1)$  of the apex of  $C$ ,

$$\text{mv}(V) = \text{conv}(k, x). \quad (19)$$

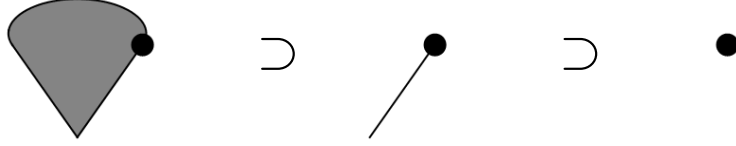


Figure 3: A non-exposed face. The drawing shows a repeated inclusion of exposed faces. The segment is an exposed face of the convex surface and the point is an exposed face of the segment. But the point is a non-exposed face of the convex surface.

**Lemma 2.4.** *Let  $V \in \mathbb{G}$  be a 2D plane. Then  $x \in k$  if and only if  $\frac{\pi}{3} \leq \varphi \leq \frac{\pi}{2}$ . If  $\varphi = 0$ , then the mean value set  $\text{mv}(V)$  is a triangle. If  $0 < \varphi < \frac{\pi}{3}$ , then  $k$  is an ellipse and the tangent points of the two tangent lines from  $x$  to  $k$  are non-exposed faces of  $\text{mv}(V)$ . If  $\frac{\pi}{3} \leq \varphi \leq \frac{\pi}{2}$ , then  $\text{mv}(V) = k$  is an ellipse.*

*Proof:* By the orbit isomorphism (11) we assume  $V \subset U$  and we use (19). Instead of the passive view of a moving plane, we imagine a rotating cone  $C$  against the horizontal plane  $V$ . The symmetry axis  $l$  of  $C$  through  $\frac{1}{2}\mathbb{1}_2 \oplus 0$  and  $0_2 \oplus 1$  follows the direction of  $z$ . We fix a vertical plane through  $l$ . This meets  $C$  in an equilateral triangle  $\Delta$ . The side  $b$  of  $\Delta$ , which is not incident with the apex  $0_2 \oplus 1$  of  $C$  runs in the base disk  $K$  of  $C$  and  $b$  is therefore pivotal to decide about  $x \in k$ . This happens if and only if  $b$  covers the other two sides of  $\Delta$ , so  $x \in k$  if and only if  $\angle(V, b) \leq \frac{\pi}{6}$ . As  $l$  and  $b$  are perpendicular, we have  $x \in k \iff \frac{\pi}{3} \leq \varphi \leq \frac{\pi}{2}$ .

At  $\varphi = 0$  the symmetry axis  $l$  of the base disk  $K$  of  $C$  is parallel to  $V$ , so  $k = \pi_V(K)$  is a segment and the mean value set  $\text{mv}(V) = \text{conv}(k, x)$  is a triangle. Otherwise, if  $\varphi > 0$ , then  $k$  is an ellipse. The case of  $\frac{\pi}{3} \leq \varphi \leq \frac{\pi}{2}$  with  $x \in k$  reduces to  $\text{mv}(V) = k$ . In the case  $0 < \varphi < \frac{\pi}{3}$  with  $x \notin k$  there are two tangent lines from  $x$  to the ellipse  $k$  with tangent points  $t_1, t_2 \in k$ . A supporting hyperplane that meets  $t_i$  must include the tangent line segment  $[x, t_i]$ , so  $\{t_i\}$  is not an exposed face of  $\text{mv}(V)$  for  $i = 1, 2$ . On the other hand  $t_i$  is an endpoint of the exposed face  $[x, t_i]$  and hence is a poonem (or a face likewise) of  $\text{mv}(V)$  for  $i = 1, 2$ . We conclude that  $\{t_1\}$  and  $\{t_2\}$  are non-exposed faces of  $\text{mv}(V)$ .  $\square$

In Figure 2 mean value sets  $\text{mv}(V)$  are drawn at the angles  $\varphi \in \{\frac{k\pi}{12} \mid k = 0, \dots, 6\}$ . The third drawing at  $\varphi = \frac{\pi}{6}$  is also used in Figure 3 to demonstrate non-exposed faces.

The results of this section show that non-exposed faces of a mean value set are typical in the following sense. A continuous curve  $\gamma : [0, 1] \rightarrow \mathbb{G}$  induces a curve of mean value sets  $\lambda \mapsto \text{mv}(\gamma(\lambda))$ . By Lemma 2.4 a mean value set without non-exposed faces must be a triangle or an ellipse. If  $\gamma$  connects the classical mean value set of a triangle to an ellipse, then we have  $\angle(\gamma(0), z) = 0$  and  $\angle(\gamma(1), z) \in [\frac{\pi}{3}, \frac{\pi}{2}]$ . Since the angle  $\varphi$  is continuous on  $\mathbb{G}$ , the curve  $\gamma$  must cross the range of angles  $(0, \frac{\pi}{3})$  with mean value sets having non-exposed faces. This range corresponds to an open subset of the Grassmannian  $\mathbb{G}$ .

### 3 Closures of exponential families

In the sequel of this article we assume that  $\mathcal{A}$  is a  $*$ -subalgebra of  $\text{Mat}(N, \mathbb{C})$  and that  $\mathcal{E}$  is an exponential family in  $\mathcal{A}$  with canonical parameter space  $\Theta$  and canonical tangent space  $V$ . The previous section has revealed for two-dimensional mean value sets in the algebra  $\text{Mat}(2, \mathbb{C}) \oplus \mathbb{C}$  that an angle of  $\varphi = \frac{\pi}{3}$  divides mean value sets with non-exposed faces from others without non-exposed faces. We show that the Staffenberg family at  $\varphi = \frac{\pi}{3}$  has a discontinuous entropy distance. We compute closures of the swallow family and of the Staffenberg family based on a few general results.

#### 3.1 General results

In this section we prove a set-theoretic upper bound for the norm closure  $\overline{\mathcal{E}}$  of the exponential family  $\mathcal{E}$ . We describe the geodesic closure of  $\mathcal{E}$  as a union of exponential families in compressed algebras.



We compute the entropy distance from e-geodesics and their limit points.

First we focus on the norm closure. There is an upper bound on the supplementary set  $\bar{\mathcal{E}} \setminus \mathcal{E}$ . This depends only on the canonical tangent space.

**Lemma 3.1.** *Let  $\mathcal{E}$  be a linear exponential family in  $\mathcal{A}$  and assume  $(\rho_i) \subset \mathcal{E}$  is a sequence of states with norm limit  $\rho = \lim_{i \rightarrow \infty} \rho_i \in \bar{\mathcal{S}}(\mathcal{A})$ . Either  $\rho \in \mathcal{E}$  or there exists a non-zero vector  $v$  in the canonical tangent space  $V$  of  $\mathcal{E}$ , such that  $\rho$  belongs to the exposed face  $F(\bar{\mathcal{S}}(\mathcal{A}), v)$  of the state space.*

*Proof:* If the sequence  $\theta_i := \ln_0(\rho_i)$  of canonical parameter values has a bounded subsequence, then this subsequence has an accumulation point  $\theta \in V = \ln_0(\mathcal{E})$  and then by continuity of  $\exp_1$  we have  $\rho = \exp_1(\theta)$ . Otherwise  $\lim_{i \rightarrow \infty} \|\theta_i\| = \infty$  and then there exists a converging subsequence (w.l.o.g. we keep  $(\theta_i)$  and  $(\rho_i)$ ), such that  $v := \lim_{i \rightarrow \infty} \frac{\theta_i}{\|\theta_i\|} \in V$ . An argument from Lemma VII in [Wi63], which can be proved by standard perturbation theory (see Corollary 6.26 in [We09]), is

$$\langle \rho, v \rangle = \lim_{i \rightarrow \infty} \langle \rho_i, v \rangle = \mu_+(v).$$

Now Lemma 2.1 proves that  $\rho \in F(\bar{\mathcal{S}}(\mathcal{A}), v)$ . □

An *e-geodesic* is a one-dimensional exponential family. The *free energy*, defined for  $a \in \mathcal{A}_{\text{sa}}$  by  $F(a) := \ln(\text{tr}(e^a))$  is a useful function to discuss the structure of limit points of e-geodesics included in an exponential family.

**Lemma 3.2.** *Suppose  $\theta, u \in \mathcal{A}_{\text{sa}}$  and  $p := p_+(u)$  is the maximal projector of  $u$ . We have*

$$\lim_{\lambda \rightarrow \infty} \exp_1(\theta + \lambda u) = \frac{p e^{p \theta p}}{\text{tr}(p e^{p \theta p})} \quad (20)$$

and

$$\lim_{\lambda \rightarrow \infty} (F(\theta + \lambda u) - \lambda \mu_+(u)) = \ln(\text{tr}(p e^{p \theta p})). \quad (21)$$

*Proof:* If  $u$  has maximal eigenvalue  $\mu_+(u) = 0$  then  $\lim_{\lambda \rightarrow \infty} e^{\theta + \lambda u} = p e^{p \theta p}$ . This equation is proved by standard perturbation theory. By the invariance for  $\alpha \in \mathbb{R}$  of  $\exp_1(\theta + \alpha \mathbb{1}) = \exp_1(\theta)$ , one obtains (20). The limit  $\lim_{\lambda \rightarrow \infty} e^{\theta + \lambda u} = p e^{p \theta p}$  implies (21) through the equivariance for  $\alpha \in \mathbb{R}$  of the free energy  $F(\theta + \alpha \mathbb{1}) = F(\theta) + \alpha$ . □

The only extension of the exponential family  $\mathcal{E}$ , that we can compute in general in this article, is the *geodesic closure*

$$\text{cl}_{\text{geo}}(\mathcal{E}) = \mathcal{E} \cup \{ \text{limit points of e-geodesics in } \mathcal{E} \}.$$

We show that  $\text{cl}_{\text{geo}}(\mathcal{E})$  is a union of exponential families. If  $p \in \mathcal{A}$  is a non-zero orthogonal projector, then the *compressed algebra* by  $p$  is

$$p\mathcal{A}p := \{p a p \mid a \in \mathcal{A}\}. \quad (22)$$

The algebra  $p\mathcal{A}p$  is a  $C^*$ -subalgebra of  $\mathcal{A}$  with identity  $p$ . The Hilbert space representation induced from  $\mathcal{A}$  makes the embedding  $p\mathcal{A}p \rightarrow \mathcal{A}$  trace-preserving. We denote functions on  $p\mathcal{A}p$  that depend on the representation by a superscript  $p$ , e.g.  $\ln^p(p) = 0$ , while  $\ln(p)$  is not defined. For  $a \in p\mathcal{A}p$  we notice  $\exp^p(a) = p \exp(a)$ ,  $\exp_1^p(a) = \frac{p e^a}{\text{tr}(p e^a)}$  and  $F^p(a) = \ln \text{tr}(p e^a)$ . The projection

$$c^p : \mathcal{A} \longrightarrow p\mathcal{A}p, \quad a \longmapsto p a p - p \frac{\text{tr}(p a)}{\text{tr}(p)}$$

is important to define the exponential family in  $p\mathcal{A}p$

$$\mathcal{E}^p := \{ \exp_1^p \circ c^p(\theta) \mid \theta \in \Theta \}.$$

This has canonical parameter space  $c^p(\Theta)$  and canonical tangent space  $c^p(V)$ .



**Proposition 3.3.** *Suppose  $\mathcal{E}$  is an exponential family in  $\mathcal{A}$ . The geodesic closure of  $\mathcal{E}$  is*

$$\text{cl}_{\text{geo}}(\mathcal{E}) = \bigcup_p \mathcal{E}^p,$$

where the disjoint union extends over the maximal projectors  $p = p_+(v)$  of all vectors  $v \in V$ .

*Proof:* This follows from (20) in Lemma 3.2.  $\square$

The *rl*-closure of the exponential family  $\mathcal{E}$  was computed by Weis [We09]. In this article we repeat this in an elementary way for the Staffenberg family and the swallow family. The *rl*-closure of  $\mathcal{E}$  is defined by

$$\text{cl}_{\text{rl}}(\mathcal{E}) = \{\rho \in \overline{\mathcal{S}}(\mathcal{A}) \mid d_{\mathcal{E}}(\rho) = 0\}.$$

We relate the geodesic closure to the *rl*-closure using asymptotics of the entropy distance  $d_{\mathcal{E}}$  along e-geodesics. By Dyson's expansion we can arrive at the derivative of the exponential, see e.g. Bhatia [Bh97]. For  $a, b \in \mathcal{A}$  this is

$$D|_a \exp(b) = \frac{\partial}{\partial x} e^{a+xb}|_{x=0} = \int_0^1 e^{ya} b e^{(1-y)a} dy.$$

Then we get for the free energy  $F$  the derivative

$$D|_a F(b) = \langle b, \exp_1(a) \rangle. \quad (23)$$

For  $\rho, \sigma \in \overline{\mathcal{S}}(\mathcal{A})$  we denote the relative entropy  $S(\rho, \sigma)$  defined in (3) by  $S_\rho(\sigma)$ . In the following proof we control the limit of  $S_\rho$  along an e-geodesic. This is remarkable, since on the subset of  $\overline{\mathcal{S}}(\mathcal{A})$ , where the relative entropy is finite, it is not continuous but only jointly lower semi-continuous (see [We78]). This happens already for the algebra  $\mathbb{C}^2$  of a bit.

**Lemma 3.4.** *Suppose  $\theta, u \in \mathcal{A}_{\text{sa}}$  and  $u$  is not proportional to the multiplicative identity  $\mathbb{1}$  in  $\mathcal{A}$ . If the state  $\rho$  belongs to the exposed face  $F(\overline{\mathcal{S}}(\mathcal{A}), u)$  of the state space, then  $S_\rho(\exp_1(\theta + \lambda u))$  is strictly monotone decreasing with  $\lambda \in \mathbb{R}$  and*

$$\inf_{\lambda \in \mathbb{R}} S_\rho(\exp_1(\theta + \lambda u)) = S_\rho\left(\lim_{\lambda \rightarrow \infty} \exp_1(\theta + \lambda u)\right).$$

*Proof:* By definition (18) of an exposed face we have for  $\tau \in \overline{\mathcal{S}}(\mathcal{A})$  the inequality  $\langle u, \tau - \rho \rangle \leq 0$  because  $\rho$  belongs to  $F(\overline{\mathcal{S}}(\mathcal{A}), u)$ . An invertible state  $\tau$  has the image  $\text{Im}(\tau) = \text{Im}(\mathbb{1}) \not\subset \text{Im}(p_+(u))$  and hence  $\tau \notin F(\overline{\mathcal{S}}(\mathcal{A}), u)$  by Lemma 2.1. This implies the strict inequality  $\langle u, \tau - \rho \rangle < 0$ , which holds for the invertible states  $\tau = \exp_1(\theta + \lambda u)$  with  $\lambda \in \mathbb{R}$ . Using (23) we have for all  $\lambda \in \mathbb{R}$

$$\frac{\partial}{\partial \lambda} S_\rho \circ \exp_1(\theta + \lambda u) = \langle u, \exp_1(\theta + \lambda u) - \rho \rangle < 0.$$

We conclude that  $S_\rho \circ \exp_1(\theta + \lambda u)$  is strictly monotone decreasing in  $\lambda$ .

We prove the infimum-limit equality. Let us consider the e-geodesic  $g : \lambda \mapsto \exp_1(\theta + \lambda u)$ . The limit for  $\lambda \rightarrow \infty$  of  $g$  is calculated in (20), we set  $p := p_+(u)$  for the maximal projector of  $u$  and

$$\sigma := \lim_{\lambda \rightarrow \infty} g(\lambda) = \exp_1^p(p\theta p).$$

The states  $\rho$  and  $\sigma$  belong to the compressed algebra  $p\mathcal{A}p$  defined in (22) and  $\sigma$  is invertible in  $p\mathcal{A}p$ . Then

$$\begin{aligned} -S(\rho, \sigma) - S(\rho) &= \text{tr}(\rho \ln^p \circ \exp_1^p(p\theta p)) = \text{tr}(\rho\theta) - F^p(p\theta p) \\ &= \lim_{\lambda \rightarrow \infty} [\text{tr}(\rho\theta) + \lambda \mu_+(u) - F(\theta + \lambda u)] \\ &= \lim_{\lambda \rightarrow \infty} [\text{tr}(\rho(\theta + \lambda u)) - F(\theta + \lambda u)] \\ &= \lim_{\lambda \rightarrow \infty} \text{tr}(\rho \ln \circ \exp_1(\theta + \lambda u)) = \lim_{\lambda \rightarrow \infty} [-S(\rho, g(\lambda)) - S(\rho)]. \end{aligned}$$

We have used (21) in the third step. The result is  $\lim_{\lambda \rightarrow \infty} S_\rho \circ g(\lambda) = S_\rho(\sigma)$ . Since  $S_\rho \circ g$  is monotone decreasing in  $\lambda$  we have  $\inf_{\lambda \in \mathbb{R}} S_\rho \circ g(\lambda) = S_\rho(\sigma)$ .  $\square$

We can relate the geodesic closure to the entropy distance and to the norm topology.

**Proposition 3.5.** Suppose  $\mathcal{E}$  is an exponential family in  $\mathcal{A}$  with canonical tangent space  $V$ . If  $v \in V$  is non-zero and  $\rho$  belongs to the exposed face  $F(\overline{\mathcal{S}}(\mathcal{A}), v)$ , then  $d_{\mathcal{E}}(\rho) = d_{\mathcal{E}^{p_+(v)}}(\rho)$ . In particular we have  $d_{\mathcal{E}}(\rho) = \inf_{\sigma \in \text{cl}_{\text{geo}}(\mathcal{E})} S(\rho, \sigma)$  and  $\text{cl}_{\text{geo}}(\mathcal{E}) \subset \text{cl}_{\text{rI}}(\mathcal{E}) \subset \overline{\mathcal{E}}$ .

*Proof:* We prove the first statement, let  $p := p_+(v)$ . If  $p_+(v) = \mathbb{1}$ , then there is nothing to prove. Otherwise  $v$  is not a multiple of  $\mathbb{1}$  and then we have by Lemma 3.4 and Lemma 3.2

$$\begin{aligned} d_{\mathcal{E}}(\rho) &= \inf_{\sigma \in \mathcal{E}} S(\rho, \sigma) = \inf_{\theta \in \Theta} \inf_{\lambda \in \mathbb{R}} S(\rho, \exp_1(\theta + \lambda v)) \\ &= \inf_{\theta \in \Theta} S(\rho, \lim_{\lambda \rightarrow \infty} \exp_1(\theta + \lambda v)) = \inf_{\theta \in \Theta} S(\rho, \exp_1^p \circ c^p(\theta)) = d_{\mathcal{E}^p}(\rho). \end{aligned}$$

Now the second statement and the inclusion  $\text{cl}_{\text{geo}}(\mathcal{E}) \subset \text{cl}_{\text{rI}}(\mathcal{E})$  follow from Proposition 3.3.

The Pinsker-Csiszár inequality, proved e.g. in the book by Petz [Pe08], says that  $\|\rho - \sigma\|_1^2 \leq \frac{1}{2} S(\rho, \sigma)$  holds for all states  $\rho, \sigma \in \overline{\mathcal{S}}(\mathcal{A})$  with the trace norm indicated by  $\|\cdot\|_1$ . This implies the inclusion of  $\text{cl}_{\text{rI}}(\mathcal{E}) \subset \overline{\mathcal{E}}$ .  $\square$

### 3.2 The Staffellberg family

The Staffellberg family  $\mathcal{E}$  corresponds to an angle of  $\varphi = \frac{\pi}{3}$  in (12), the special point in the metamorphosis of Figure 2. We calculate closures of  $\mathcal{E}$  and we prove  $\text{cl}_{\text{geo}}(\mathcal{E}) = \text{cl}_{\text{rI}}(\mathcal{E}) \subsetneq \overline{\mathcal{E}}$ . This implies that the entropy distance from  $\mathcal{E}$  has a discontinuity.

**Definition 3.6.** The *Staffellberg*<sup>8</sup> family is the linear exponential family in  $\mathcal{A} := \text{Mat}(2, \mathbb{C}) \oplus \mathbb{C}$  defined by

$$\mathcal{E} := \exp_1(\text{span}_{\mathbb{R}}\{\sigma_1 \oplus 0, \sigma_2 \oplus 1\}).$$

The canonical tangent space  $V = \Theta$  of  $\mathcal{E}$  is spanned by  $v_1 := \sigma_1 \oplus 0$  and  $v_2 := \sigma_2 \oplus 1 - \frac{1}{3}\mathbb{1}$ . The vector  $z = -\frac{1}{2}\mathbb{1}_2 \oplus 1$  is perpendicular to  $v_1$ , so

$$\varphi = \angle(V, z) = \angle(v_2, z) = \arccos(\frac{1}{2}) = \frac{\pi}{3}.$$

The Staffellberg family  $\mathcal{E}$  is included in the 3D cone  $C$  introduced in (14). We use the notation  $B := \{\rho(\alpha) \mid \alpha \in (0, 2\pi)\}$  for the *punctured base circle* of  $C$  with  $\rho(0) = \frac{1}{2}(\mathbb{1}_2 + \sigma_2) \oplus 0$  missing. In Corollary 3.9 we prove that  $d_{\mathcal{E}}$  is discontinuous at  $\rho(0)$  because  $B \subset \text{cl}_{\text{rI}}(\mathcal{E})$  while  $d_{\mathcal{E}}(\rho(0)) = \ln(2)$ .

The symmetry axis  $l$  of  $C$  goes through the tracial state  $\frac{1}{3}\mathbb{1}$  and the apex  $0_2 \oplus 1$ , where the generating lines of  $C$  meet with  $l$  under an angle of  $\frac{\pi}{6}$ . The generating line  $[\rho(0), 0_2 \oplus 1]$  is perpendicular to  $V$ . We denote by  $c := \frac{1}{2}(\rho(0) + 0_2 \oplus 1)$  the midpoint. The basis vectors of  $V$  connect special points in  $C$ ,

$$v_1 = \sigma_1 \oplus 0 = \rho(\frac{\pi}{2}) - \rho(\frac{3}{2}\pi) \quad \text{and} \quad v_2 = \sigma_2 \oplus 1 - \frac{1}{3}\mathbb{1} = \frac{4}{3}(c - \rho(\pi)).$$

The algebra generated by  $\sigma_2 \oplus 1$  is isomorphic to  $\mathbb{C}^2$  and it has the segment  $[\rho(\pi), c]$  as its state space. The e-geodesic  $\{\exp_1(\lambda(\sigma_2 \oplus 1)) \mid \lambda \in \mathbb{R}\}$  is included in  $\mathcal{E}$  and it covers the invertible states in  $[\rho(\pi), c]$ . The Staffellberg family is depicted in Figure 4.

**Remark 3.7.** Let us discuss a polar parametrization of the Staffellberg family  $\mathcal{E}$ . The zero trace of  $v_1$  and  $v_2$  above is not important. We use

$$u(\alpha) := \sin(\alpha)\sigma_1 \oplus 0 + \cos(\alpha)(\sigma_2 \oplus 1), \quad \alpha \in \mathbb{R}. \quad (24)$$

We need for non-zero  $b \in \mathbb{R}^3$  and  $|b| := \sqrt{b_1^2 + b_2^2 + b_3^2}$  the formula in  $\text{Mat}(2, \mathbb{C})$

$$\exp(b\hat{\sigma}) = \cosh(|b|)\mathbb{1}_2 + \sinh(|b|)\frac{b}{|b|}\hat{\sigma}.$$

Writing  $b(\alpha) = (\sin(\alpha), \cos(\alpha), 0)$  we parametrize  $\mathcal{E}$  for  $\alpha \in \mathbb{R}$  and  $t \geq 0$  by

$$\sigma(\alpha, t) := \exp_1(u(\alpha)t) = \frac{1}{T(\alpha, t)} \left[ (\mathbb{1}_2 \cosh(t) + b(\alpha)\hat{\sigma} \sinh(t)) \oplus e^{\cos(\alpha)t} \right]. \quad (25)$$

<sup>8</sup>That exponential family has the form of the *Staffellberg* mountain, in the natural preserve of *Veldensteiner Forst*.

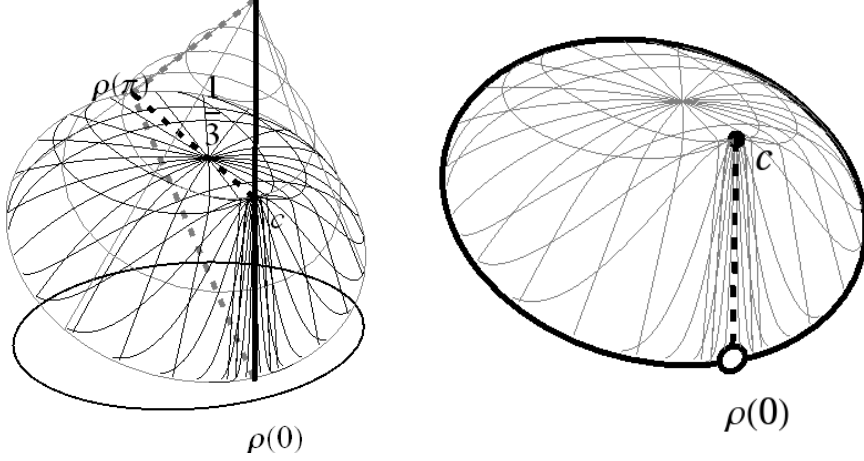


Figure 4: An e-geodesic model of the Staffellberg family  $\mathcal{E}$ . Left: The elliptic mean value set below indicates the alignment of the canonical tangent space  $V$ . The cone  $C$  containing  $\mathcal{E}$  has apex  $0_2 \oplus 1$  and the generating line  $[\rho(0), 0_2 \oplus 1]$  (black, thick) is perpendicular to  $V$ . The open segment (black, dashed) from the midpoint  $c$  of this generating line through  $\frac{1}{3}\mathbb{1}$  to  $\rho(\pi)$  belongs to  $\mathcal{E}$ . The large equilateral triangle is a section through the symmetry axis of  $C$ , it is isometrically isomorphic to the state space of  $\mathbb{C}^3$ . Right: The geodesic closure  $\text{cl}_{\text{geo}}(\mathcal{E})$  covers the punctured base circle of  $C$  (large circle) with  $\rho(0)$  (small circle) missing. The geodesic closure also includes  $c$  (thick). The segment  $[\rho(0), c]$  (dashed) belongs to the norm closure  $\bar{\mathcal{E}}$ .

Here  $T(\alpha, t) := 2 \cosh(t) + e^{\cos(\alpha)t}$  is for normalization. The vectors  $v_1$  and  $v_2$  are completed by  $v_3 := -\rho(0) \oplus 1$  to an orthogonal basis of  $U = V + \mathbb{R}z$ . We obtain for all  $\alpha \in \mathbb{R}$  and  $t \geq 0$

$$\langle \sigma(\alpha, t), v_3 \rangle = \frac{1}{T(\alpha, t)} (-\sinh(t) \cos(\alpha) - \cosh(t) + e^{t \cos(\alpha)}). \quad (26)$$

We discuss closures of the Staffellberg family and its entropy distance.

**Theorem 3.8.** *The Staffellberg family  $\mathcal{E}$  has geodesic closure and rl-closure equal to  $\text{cl}_{\text{geo}}(\mathcal{E}) = \text{cl}_{\text{rl}}(\mathcal{E}) = \mathcal{E} \cup B \cup \{c\}$ , the norm closure is  $\bar{\mathcal{E}} = \mathcal{E} \cup B \cup [\rho(0), c]$ . The entropy distance of  $\rho \in [\rho(0), 0_2 \oplus 1]$  from  $\mathcal{E}$  is  $d_{\mathcal{E}}(\rho) = S(\rho, c)$ .*

*Proof:* We calculate the geodesic closure  $\text{cl}_{\text{geo}}(\mathcal{E})$  using Proposition 3.3 as a union of exponential families  $\mathcal{E}^q = \{\exp_1^q \circ c^q(\theta) \mid \theta \in V\}$  for maximal projectors  $q$ . In place of the maximal projectors of  $v \neq 0$  in  $V$  we consider the maximal projectors of the vectors  $\{u(\alpha) \mid \alpha \in \mathbb{R}\}$  in (24). We distinguish two cases depending on the spectral projectors in the orthogonal sum

$$u(\alpha) = \rho(\alpha) - \rho(\alpha + \pi) + 0_2 \oplus \cos(\alpha).$$

The maximal eigenvalue of  $u(\alpha)$  is constant one. If  $\alpha \neq 0 \bmod 2\pi$ , then the maximal projector of  $u(\alpha)$  is  $\rho(\alpha)$ . This has rank one and  $\mathcal{E}^{\rho(\alpha)} = \{\rho(\alpha)\}$ . We have proved  $B \subset \text{cl}_{\text{geo}}(\mathcal{E})$ . If  $\alpha = 0 \bmod 2\pi$ , then the maximal projector of  $u(0)$  is  $p := \rho(0) + 0_2 \oplus 1 = 2c$ . The projections  $p(\sigma_1 \oplus 0)p = 0$  and  $p(\sigma_2 \oplus 1)p = p$  show that the canonical parameter space of the exponential family  $\mathcal{E}^p$  is  $c^p(\Theta) = c^p(V) = \{0\}$ . We get  $\mathcal{E}^p = \{c\}$  and we conclude  $\text{cl}_{\text{geo}}(\mathcal{E}) = \mathcal{E} \cup B \cup \{c\}$ .

For a discussion of the norm closure we consider exposed faces as proposed in Lemma 3.1. This provides an upper bound on  $\bar{\mathcal{E}}$  in terms of exposed faces by vectors  $v \in V$ . These faces depend only on the maximal projector of  $v$  by Lemma 2.1, so we have

$$\bar{\mathcal{E}} \subset \mathcal{E} \cup \bigcup_{\alpha \in \mathbb{R}} F(\bar{\mathcal{S}}(\mathcal{A}), u(\alpha)).$$

We start with a one-dimensional face. The segment  $[\rho(0), 0_2 \oplus 1]$  is the exposed face  $F(\bar{\mathcal{S}}(\mathcal{A}), u(0))$  and we have  $p_+(u(0)) = p$ . The fact  $\mathcal{E}^p = \{c\}$  and Proposition 3.5 show for  $\rho \in [\rho(0), 0_2 \oplus 1]$

$$d_{\mathcal{E}}(\rho) = d_{\mathcal{E}^p}(\rho) = S(\rho, c). \quad (27)$$

For  $\alpha \neq 0 \bmod 2\pi$  the maximal projector  $\rho(\alpha)$  of  $u(\alpha)$  has rank one and the exposed face is  $F(\overline{\mathcal{S}}(\mathcal{A}), u(\alpha)) = \{\rho(\alpha)\}$ . We obtain

$$\overline{\mathcal{E}} \subset \mathcal{E} \cup B \cup [\rho(0), 0_2 \oplus 1].$$

Since  $B \subset \text{cl}_{\text{geo}}(\mathcal{E})$  Proposition 3.5 shows  $B \subset \overline{\mathcal{E}}$ . We finish the calculation of  $\overline{\mathcal{E}}$  by showing that exactly the half  $[\rho(0), c]$  of the segment  $[\rho(0), 0_2 \oplus 1]$  belongs to  $\overline{\mathcal{E}}$ .

We prove that at most the half segment  $[\rho(0), c]$  belongs to  $\overline{\mathcal{E}}$  by showing that  $\mathcal{E}$  lies completely in one of the closed half spaces of  $\frac{1}{3}\mathbb{1} + U$  divided by the hyperplane  $c + V$ . This works because the vector  $v_3$  defined in the paragraph after (25) is orthogonal to  $V$  and

$$\langle \rho(0), v_3 \rangle = -1, \quad \langle c, v_3 \rangle = 0, \quad \text{and} \quad \langle 0_2 \oplus 1, v_3 \rangle = 1.$$

It remains to prove  $\langle \sigma, v_3 \rangle \leq 0$  for all states  $\sigma$  in  $\mathcal{E}$ . In the polar parametrization (25) of  $\mathcal{E}$ ,  $\mathbb{R} \times \mathbb{R}_0^+ \rightarrow \mathcal{E}$ ,  $(\alpha, t) \mapsto \sigma(\alpha, t)$  the denominator  $T(\alpha, t)$  of  $\langle \sigma(\alpha, t), v_3 \rangle$  is strictly positive (26). So we can prove for  $\alpha \in \mathbb{R}$ ,  $t \geq 0$  and

$$z(\alpha, t) := T(\alpha, t) \langle \sigma(\alpha, t), v_3 \rangle = -\cos(\alpha) \sinh(t) - \cosh(t) + e^{\cos(\alpha)t},$$

that  $z(\alpha, t) \leq 0$  holds. For all  $\alpha \in \mathbb{R}$  we have  $z(\alpha, 0) = 0$  and for all  $\alpha \in \mathbb{R}$  and  $t \geq 0$

$$\frac{\partial}{\partial t} z(\alpha, t) \pm z(\alpha, t) = (\cos(\alpha) \pm 1) [e^{\cos(\alpha)t} - e^{\pm t}] \leq 0$$

implies  $\frac{\partial}{\partial t} z(\alpha, t) \leq 0$  and by integration from 0 to  $t$  we get  $z(\alpha, t) \leq 0$ .

We show  $[\rho(0), c] \subset \overline{\mathcal{E}}$ . The state  $\rho(0)$  lies in the closure of  $B$  and  $c$  is the limit of an e-geodesic in  $\mathcal{E}$ . We still have to approximate for  $\lambda \in (0, 1)$  the state  $\tau(\lambda) := (1 - \frac{\lambda}{2})\rho(0) \oplus \frac{\lambda}{2}$  from within  $\mathcal{E}$ . For  $t > 0$  we choose  $\alpha(t) := \sqrt{\frac{2}{t} \ln(\frac{2-\lambda}{\lambda})}$  with  $\lim_{t \rightarrow \infty} \alpha(t) = 0$  and  $\lim_{t \rightarrow \infty} e^{(\cos(\alpha(t))-1)t} = \frac{\lambda}{2-\lambda}$ . Then we expand numerator and denominator of  $\sigma(\alpha(t), t)$  defined in (25) by the factor  $e^{-t}$  and we get

$$\lim_{t \rightarrow \infty} \sigma(\alpha(t), t) = \frac{\frac{1}{2}(\mathbb{1}_2 + \sigma_2) \oplus \frac{\lambda}{2-\lambda}}{1 + \frac{\lambda}{2-\lambda}} = \tau(\lambda).$$

Finally we calculate the rl-closure. This is sandwiched by Proposition 3.5 between geodesic and norm closures

$$\mathcal{E} \cup B \cup \{c\} \subset \text{cl}_{\text{rl}}(\mathcal{E}) \subset \mathcal{E} \cup B \cup [\rho(0), c].$$

It remains to discuss states  $\rho \in [\rho(0), 0_2 \oplus 1]$ . We use  $d_{\mathcal{E}}(\rho) = S(\rho, c)$  proved in (27), whence  $d_{\mathcal{E}}(\rho) = 0$  implies  $\rho = c$ .  $\square$

**Corollary 3.9.** *The entropy distance  $d_{\mathcal{E}}$  from the Staffenberg family is discontinuous at  $\rho(0)$ .*

*Proof:* By the previous theorem we have  $d_{\mathcal{E}}(\rho(0)) = S(\rho(0), c) = \ln(2)$  while  $d_{\mathcal{E}} \equiv 0$  on the punctured base circle  $B$  of the cone  $C$ . This includes  $\rho(0)$  in its norm closure.  $\square$

### 3.3 The swallow family

We now consider 2D families  $\mathcal{E} = \exp_1(V)$  that have non-exposed faces in the mean value set. By Lemma 2.4 this corresponds to an angle  $\varphi(V) \in (0, \pi/3)$ , see (12). We calculate closures of  $\mathcal{E}$  and we prove  $\text{cl}_{\text{geo}}(\mathcal{E}) \subsetneq \text{cl}_{\text{rl}}(\mathcal{E}) = \overline{\mathcal{E}}$ .

Calculations become easy for  $\varphi = \arccos(\sqrt{\frac{2}{5}}) \approx 0.28\pi$  and we then call  $\mathcal{E}$  the *swallow family*:

**Definition 3.10.** The *swallow family* is the linear exponential family in  $\mathcal{A} := \text{Mat}(2, \mathbb{C}) \oplus \mathbb{C}$  defined by

$$\mathcal{E} := \exp_1(\text{span}_{\mathbb{R}}\{\sigma_1 \oplus 1, \sigma_2 \oplus 1\}).$$

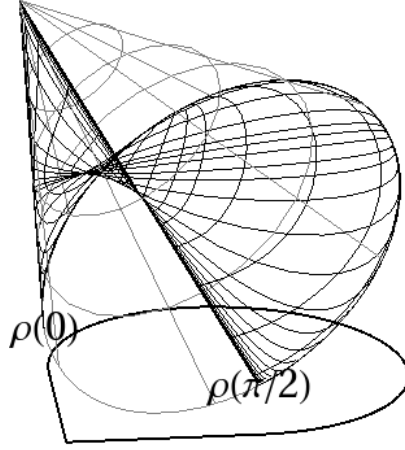


Figure 5: An e-geodesic model of the swallow family  $\mathcal{E}$ . The mean value set below  $\mathcal{E}$  indicates the alignment of the canonical tangent space  $V$ . The cone  $C$  containing  $\mathcal{E}$  has apex  $0_2 \oplus 1$  and the generating lines  $[\rho(0), 0_2 \oplus 1]$  and  $[\rho(\frac{\pi}{2}), 0_2 \oplus 1]$  (black) belong to the rl-closure of  $\mathcal{E}$ . Their lower corners  $\rho(0)$  and  $\rho(\frac{\pi}{2})$  do not belong to the geodesic closure of  $\mathcal{E}$ , they project to the non-exposed faces of the mean value set.

The canonical tangent space  $V = \Theta$  of  $\mathcal{E}$  is spanned by  $v_1 := \sigma_1 \oplus 1 - \frac{1}{3}\mathbb{1}$  and  $v_2 := \sigma_2 \oplus 1 - \frac{1}{3}\mathbb{1}$ . The vector  $z = -\frac{1}{2}\mathbb{1}_2 \oplus 1$  is perpendicular to  $v_2 - v_1$ , so indeed

$$\varphi = \angle(V, z) = \angle(v_1 + v_2, z) = \arccos(\sqrt{\frac{2}{5}}).$$

In Lemma 2.4 we have seen that this angle produces non-exposed faces at the mean value set  $\text{mv}(V)$ . The swallow family is included in the 3D cone  $C$  defined in (14). We have  $\text{mv}(V) = \pi_V(C)$  and the points  $\rho(0) = \frac{1}{2}(\mathbb{1}_2 + \sigma_2) \oplus 0$  and  $\rho(\frac{\pi}{2}) = \frac{1}{2}(\mathbb{1}_2 + \sigma_1) \oplus 0$  on the base circle of  $C$  project to the non-exposed faces  $\pi_V(\rho(0))$  and  $\pi_V(\rho(\frac{\pi}{2}))$  of the mean value set, see Proposition 3.11. They are also significant because we prove in Theorem 3.12 that they are missing in the geodesic closure  $\text{cl}_{\text{geo}}(\mathcal{E})$  while they belong to rl-closure  $\text{cl}_{\text{rl}}(\mathcal{E})$ . The swallow family is depicted in Figure 5.

**Proposition 3.11.** *If  $V$  denotes the canonical tangent space of the swallow family, then  $\{\pi_V(\rho(0))\}$  and  $\{\pi_V(\rho(\frac{\pi}{2}))\}$  are non-exposed faces of the mean value set  $\text{mv}(V)$ .*

*Proof:* The mean value set is the projection of the cone  $C$  to  $V$ . It is the convex hull of the ellipse  $k$ , which is the projected base disk of  $C$ , and of the point  $x = \pi_V(0_2 \oplus 1)$ , which is the projected apex, see (19). Let us denote the boundary of  $k$  by  $e$ . Since  $\varphi \approx 0.28\pi < \frac{\pi}{3}$ , we have  $x \notin k$  and the tangent points  $t_1$  and  $t_2$  from  $x$  to  $e$  are non-exposed faces of  $\text{mv}(V)$  by Lemma 2.4. We calculate  $t_1$  and  $t_2$  using the polar of  $x$  with respect to the conic  $e$  (in the sense of projective geometry): we have to find on the real space  $V$  a symmetric bilinear form  $\beta$  including linear and constant terms, such that  $e = \{v \in V \mid \beta(v, v) = 0\}$ . Then the *polar*  $\{v \in V \mid \beta(v, x) = 0\}$  of  $x$  intersects  $e$  in  $t_1$  and  $t_2$ , see e.g. Section 3.5.11 in Fischer [Fi85].

The ellipse  $e$  consists of the projected base circle points  $\{\pi_V(\rho(\alpha)) \mid \alpha \in \mathbb{R}\}$  of the cone  $C$ . We can neglect the projection to  $V$  by using the linear forms  $\eta : \mathcal{A}_{\text{sa}} \rightarrow \mathbb{R}$ ,  $a \mapsto \langle a, v_1 \rangle$  and  $\xi : \mathcal{A}_{\text{sa}} \rightarrow \mathbb{R}$ ,  $a \mapsto \langle a, v_2 \rangle$ , which are invariant under  $\pi_V$ . These evaluate for  $\alpha \in \mathbb{R}$  as

$$\eta(\rho(\alpha)) = \sin(\alpha) - \frac{1}{3} \quad \text{and} \quad \xi(\rho(\alpha)) = \cos(\alpha) - \frac{1}{3}.$$

The bilinear form

$$\beta : \mathcal{A}_{\text{sa}} \times \mathcal{A}_{\text{sa}} \longrightarrow \mathbb{R}, \quad (a, b) \longmapsto \eta(a)\eta(b) + \xi(a)\xi(b) + \frac{1}{3}(\eta + \xi)(a + b) - \frac{7}{9}$$

satisfies  $\beta(\rho(\alpha), \rho(\alpha)) = 0$  for all  $\alpha \in \mathbb{R}$  and it satisfies  $\beta(\rho(0), 0_2 \oplus 1) = \beta(\rho(\frac{\pi}{2}), 0_2 \oplus 1) = 0$ .  $\square$

We denote the *open segment* between  $a, b \in \mathcal{A}_{\text{sa}}$  by  $]a, b[ := \{(1 - \lambda)a + \lambda b \mid \lambda \in (0, 1)\}$ .

**Theorem 3.12.** *The geodesic closure of the swallow family  $\mathcal{E}$  is the union of  $\mathcal{E}$  with the rank-two states in the segments  $]\rho(0), 0_2 \oplus 1[$  and  $]\rho(\frac{\pi}{2}), 0_2 \oplus 1[$  and with the pure states  $0_2 \oplus 1$  and  $\{\rho(\alpha) \mid \frac{\pi}{2} < \alpha < 2\pi\}$ . The rl- and norm closures are  $\text{cl}_{\text{rl}}(\mathcal{E}) = \bar{\mathcal{E}} = \text{cl}_{\text{geo}}(\mathcal{E}) \cup \{\rho(0), \rho(\frac{\pi}{2})\}$ .*

*Proof:* We repeat some arguments from Theorem 3.8. First we calculate the geodesic closure  $\text{cl}_{\text{geo}}(\mathcal{E})$  using Proposition 3.3. For  $\alpha \in \mathbb{R}$  we have the orthogonal sum

$$u(\alpha) := \sin(\alpha)(\sigma_1 \oplus 1) + \cos(\alpha)(\sigma_2 \oplus 1) = \rho(\alpha) - \rho(\alpha + \pi) + 0_2 \oplus \sqrt{2} \cos(\alpha - \frac{\pi}{4}).$$

The maximal projectors are

$$p := p_+(u(0)) = \rho(0) + 0_2 \oplus 1, \quad q := p_+(u(\frac{\pi}{2})) = \rho(\frac{\pi}{2}) + 0_2 \oplus 1,$$

$p_+(u(\alpha)) = 0_2 \oplus 1$  for  $0 < \alpha < \frac{\pi}{2}$  and  $p_+(u(\alpha)) = \rho(\alpha)$  for  $\frac{\pi}{2} < \alpha < 2\pi$ . We notice that  $p(\sigma_1 \oplus 1)p = 0_2 \oplus 1$  and this implies that the projection  $c^p : \mathcal{A} \rightarrow p\mathcal{A}p \cong \mathbb{C}^2$  restricted to  $V$  has the image  $\mathbb{R}(0_2 \oplus 1 - \rho(0)) \cong \mathbb{R}(1, -1) \subset \mathbb{C}^2$ . The analogue arguments apply to  $q$ , so the exponential family

$$\mathcal{E}^p = ]\rho(0), 0_2 \oplus 1[ \quad \text{resp.} \quad \mathcal{E}^q = ]\rho(\frac{\pi}{2}), 0_2 \oplus 1[$$

in  $p\mathcal{A}p$  resp.  $q\mathcal{A}q$  consists of the invertible states in the compressed algebra  $p\mathcal{A}p$  resp.  $q\mathcal{A}q$ . All other maximal projectors  $r$  of elements of  $v \neq 0$  of  $V$  have rank one and produce the exponential family  $\mathcal{E}^r = \{\exp_1^r \circ c^r(\theta) \mid \theta \in V\} = \{r\}$ . This completes the calculation of the geodesic closure.

In the second step we prove that the points  $\rho(0)$  and  $\rho(\frac{\pi}{2})$  missing in the geodesics closure belong to the rl-closure of the swallow family. By Lemma 2.1 we have the exposed face  $F(\bar{\mathcal{S}}(\mathcal{A}), u(0)) = [\rho(0), 0_2 \oplus 1]$ , which contains the state  $\rho(0)$ . Then we obtain from Proposition 3.5

$$d_{\mathcal{E}}(\rho(0)) = d_{\mathcal{E}^p}(\rho(0)).$$

We may use the e-geodesic  $g : \lambda \mapsto \exp_1^p(\lambda\rho(0))$  to prove that the state  $\rho(0)$  lies in  $\text{cl}_{\text{geo}}(\mathcal{E}^p)$ . So  $d_{\mathcal{E}^p}(\rho(0)) = 0$  holds and this implies  $d_{\mathcal{E}}(\rho(0)) = 0$  and  $\rho(0) \in \text{cl}_{\text{rl}}(\mathcal{E})$ . The analogue arguments show  $\rho(\frac{\pi}{2}) \in \text{cl}_{\text{rl}}(\mathcal{E})$ .

In the last step we use Lemma 3.1 which says that the norm closure  $\bar{\mathcal{E}}$  is included in the union of  $\mathcal{E}$  and the exposed faces of the state space  $\bar{\mathcal{S}}(\mathcal{A})$  by vectors  $v$  in  $V$ . These face are shown in Lemma 2.1 to depend only on the maximal projector of  $v$ , so we use  $u(\alpha)$  instead. The exposed faces  $F(\bar{\mathcal{S}}(\mathcal{A}), u(\alpha))$  are discussed above and we have  $\bar{\mathcal{E}} = \text{cl}_{\text{rl}}(\mathcal{E})$ .  $\square$

**Remark 3.13.** • One can establish  $\text{cl}_{\text{rl}}(\mathcal{E}) = \bar{\mathcal{E}}$  for any 2D exponential family  $\mathcal{E}$  in  $\text{Mat}(2, \mathbb{C}) \oplus \mathbb{C}$  with canonical tangent space  $V$  and  $\varphi = \angle(V, z) \neq \frac{\pi}{3}$ . This is proved by arguments analogous to Theorem 3.12 if  $\varphi < \frac{\pi}{3}$  and becomes even simpler for  $\varphi > \frac{\pi}{3}$  where  $\bar{\mathcal{E}} \setminus \mathcal{E}$  is the base circle  $\{\rho(\alpha) \mid \alpha \in [0, 2\pi)\}$  of the cone  $C$ . So the discussion following (2) shows that the Staffelfberg family is indeed the unique exponential family in the metamorphosis of Figure 2 with a discontinuous entropy distance.

- There is no e-geodesic in the swallow family  $\mathcal{E}$  that meets  $\rho(0)$  asymptotically. Calculation of  $\text{cl}_{\text{rl}}(\mathcal{E})$  in Theorem 3.12 is done by two limits of e-geodesics. One is implicit in the equation  $d_{\mathcal{E}}(\rho(0)) = d_{\mathcal{E}^p}(\rho(0))$  where e-geodesics in  $\mathcal{E}$  with limit on  $\mathcal{E}^p = ]\rho(0), 0_2 \oplus 1[$  were used. Only the second geodesic  $g$  in  $\mathcal{E}^p$  meets  $\rho(0)$  asymptotically. In the projection to  $V$  this two-stage analysis corresponds to a description of the poonem  $\pi_V(\rho(0))$  of the mean value set  $\text{mv}(V) = \pi_V(C)$  by the exposed faces  $\pi_V(C) \supset \pi_V([\rho(0), 0_2 \oplus 1]) \supset \pi_V(\rho(0))$ , see Figure 3.

## 4 Maximizers of the entropy distance

In applications to neurosciences and statistical physics it is desirable to know the local maximizers of entropy distance  $d_{\mathcal{E}}$  from the exponential family  $\mathcal{E}$ , see e.g. Ay and Knauf [AK06]. We can show that a local maximizer  $\rho$  of  $d_{\mathcal{E}}$  carries a clear imprint from its projection  $\pi_{\mathcal{E}}(\rho)$  to  $\mathcal{E}$ . This generalizes the

classical case, where  $\rho$  is the conditional probability distribution of  $\pi_{\mathcal{E}}(\rho)$  conditioned on its own support  $\text{supp}(\rho)$

$$\rho = \pi_{\mathcal{E}}(\rho)(\cdot | \text{supp}(\rho)). \quad (28)$$

We use differential calculus in a matrix algebra (see e.g. Bhatia [Bh97] for introductory literature) and compute directional derivatives of  $d_{\mathcal{E}}$  at a state along all the directions of its face. We obtain a quantum analogue of (28).

**Remark 4.1.** In the probabilistic case the assertion (28) was proved for a local maximizer  $\rho \in \mathcal{E} + V^{\perp}$  of  $d_{\mathcal{E}}$  by Ay [Ay02]. All (one-sided) directional derivatives of  $d_{\mathcal{E}}$  were computed by Matúš [Ma07] in the probabilistic case. Beyond this more difficult question, Section 5 in the latter reference contains further characterizations of a local maximizer that wait to be examined in the quantum case.

The *support projector* of a state  $\rho \in \overline{\mathcal{S}}(\mathcal{A})$  is the orthogonal projector  $p \in \mathcal{A}$  with the same image  $\text{Im}(p) = \text{Im}(\rho)$ . The von Neumann entropy of  $\rho$  is  $S(\rho) := -\text{tr}(\rho \ln(\rho))$ . We denote functions in  $p\mathcal{A}p$  by a superscript, see the paragraph following (22).

**Theorem 4.2.** Suppose  $\mathcal{A}$  is a  $*$ -subalgebra of  $\text{Mat}(N, \mathbb{C})$  and  $\mathcal{E}$  is a linear exponential family in  $\mathcal{A}$ . Let  $\rho \in \mathcal{E} + V^{\perp}$  be a state in  $\mathcal{A}$ , let  $p$  denote the support projector of  $\rho$  and put  $\theta := \ln_0 \circ \pi_{\mathcal{E}}(\rho) \in V$ . If  $u$  is a traceless self-adjoint matrix in  $p\mathcal{A}p$ , then  $D|_{\rho} d_{\mathcal{E}}(u) = \langle u, \ln^p(\rho) - \theta \rangle$ . If  $\rho$  is a local maximizer of  $d_{\mathcal{E}}$ , then  $\rho = \exp_1^p(p\theta p)$  and  $d_{\mathcal{E}}(\rho) = F(\theta) - F^p(p\theta p)$ .

*Proof:* We have discussed in the paragraph following (9) that the projection  $\pi_{\mathcal{E}}$  defined for  $a \in \mathcal{E} + V^{\perp}$  by intersection  $a \mapsto (a + V^{\perp}) \cap \mathcal{E}$  is real analytic. We obtain the real analytic mapping

$$L : \mathcal{E} + V^{\perp} \longrightarrow V, \quad a \longmapsto \ln_0 \circ \pi_{\mathcal{E}}(a).$$

The projection of  $a$  is written  $\pi_{\mathcal{E}}(a) = \exp_1 \circ L(a)$ . It can be used to describe the entropy distance (5) of a state  $\rho \in \mathcal{E} + V^{\perp}$  from  $\mathcal{E}$  by

$$\begin{aligned} d_{\mathcal{E}}(\rho) &= S(\rho, \pi_{\mathcal{E}}(\rho)) = S(\rho, \exp_1 \circ L(\rho)) \\ &= -S(\rho) - \text{tr}(\rho \ln \circ \exp_1 \circ L(\rho)) = -S(\rho) - \text{tr}(\rho L(\rho)) + F \circ L(\rho) \end{aligned} \quad (29)$$

with the free energy  $F$ . As the state  $\rho$  is positive and invertible in the algebra  $p\mathcal{A}p$ , we can differentiate at  $\rho$  the logarithm  $\ln^p$  in  $p\mathcal{A}p$  in the direction of any self-adjoint matrix  $u \in p\mathcal{A}p$ . A formula in [Li73] reads

$$D|_{\rho} \ln^p(u) = \int_0^{\infty} (\rho + sp)^{-1} u (\rho + sp)^{-1} ds$$

and we get by reordering under the trace

$$D|_{\rho} S(u) = -\langle u, \ln^p(\rho) \rangle - \text{tr}(u).$$

Using the derivative of the free energy (23), which is for  $a, b \in \mathcal{A}$  given by  $D|_a F(b) = \langle b, \exp_1(a) \rangle$ , the chain rule leads to

$$\begin{aligned} D|_{\rho}(F \circ L)(u) &= D|_{L(\rho)} F \circ D|_{\rho} L(u) \\ &= \langle D|_{\rho} L(u), \exp_1 \circ L(\rho) \rangle = \langle D|_{\rho} L(u), \pi_{\mathcal{E}}(\rho) \rangle. \end{aligned}$$

Since the image of  $L$  is  $V$  we have  $D|_{\rho} L(u) \in V$  and thus by definition of the projection  $\pi_{\mathcal{E}}$  follows  $\langle D|_{\rho} L(u), \pi_{\mathcal{E}}(\rho) - \rho \rangle = 0$ . Differentiation of (29) in the direction of a traceless self-adjoint matrix  $u \in p\mathcal{A}p$  gives

$$\begin{aligned} D|_{\rho} d_{\mathcal{E}}(u) &= \langle u, \ln^p(\rho) \rangle + \text{tr}(u) - \langle u, L(\rho) \rangle - \langle \rho, D|_{\rho} L(u) \rangle \\ &+ \langle D|_{\rho} L(u), \pi_{\mathcal{E}}(\rho) \rangle = \langle u, \ln^p(\rho) - L(\rho) \rangle. \end{aligned}$$

This completes the asserted directional derivative.

If  $\rho$  is a local maximizer of  $d_{\mathcal{E}}$ , then  $\ln^p(\rho) = pL(\rho)p + \lambda p$  for some real  $\lambda$  because  $p$  spans the orthogonal complement of the space of traceless self-adjoint matrices in  $p\mathcal{A}p$ . It follows that  $\rho$  must



be proportional to  $p \exp(p L(\rho)p)$  as claimed. If we write  $\theta := L(\rho) = \ln_0 \circ \pi_{\mathcal{E}}(\rho)$ , then we have  $\rho = \exp_1^p(p \theta p)$  and  $\pi_{\mathcal{E}}(\rho) = \exp_1(\theta)$ . We get

$$\begin{aligned} d_{\mathcal{E}}(\rho) &= S(\rho, \pi_{\mathcal{E}}(\rho)) = \text{tr}[\rho(\ln^p(\rho) - \ln \circ \pi_{\mathcal{E}}(\rho))] \\ &= \text{tr}[\rho(p \theta p - p \ln \circ \text{tr} \circ \exp^p(p \theta p) - \theta + \mathbb{1} \ln \circ \text{tr} \circ \exp(\theta))] \\ &= \ln(\text{tr}(e^{\theta})) - \ln(\text{tr}(p e^{p \theta p})). \end{aligned} \quad \square$$

**Acknowledgment:** SW thanks the organizers of the DFG research group “Geometry and Complexity in Information Theory” (2004–2008) for the great time spent at the meetings.

## References

- [AS01] Alfsen, E. M., Shultz, F. W.: *State Spaces of Operator Algebras*. Birkhäuser, Boston (2001)
- [AN00] Amari, S., Nagaoka, H.: *Methods of Information Geometry*. Translations of Mathematical Monographs 191, AMS, Providence (2000)
- [Ay02] Ay, N.: *An information-geometric approach to a theory of pragmatic structuring*. Ann. Probab. 30, 416–436 (2002)
- [AK06] Ay, N., Knauf, A.: *Maximizing multi-information*. Kybernetika 42, 517–538 (2006)
- [Ba78] Barndorff-Nielsen, O.: *Information and Exponential Families in Statistical Theory*. John Wiley & Sons, New York (1978)
- [Bh97] Bhatia, R.: *Matrix Analysis*. Springer-Verlag, New York (1997)
- [Ce82] Čencov, N. N.: *Statistical Decision Rules and Optimal Inference*. Translations of Mathematical Monographs 53, AMS, Providence (1982)
- [CM03] Csiszár, I., Matúš, F.: *Information projections revisited*. IEEE Trans. Inf. Theory 49, 1474–1490 (2003)
- [CM05] Csiszár, I., Matúš, F.: *Closures of exponential families*. Ann. Probab. 33, 582–600 (2005)
- [Fi85] Fischer, G.: *Analytische Geometrie*. Vieweg Studium 35, Vieweg, Braunschweig, 4th ed. (1985)
- [Gr67] Grünbaum, B.: *Convex Polytopes*. Springer-Verlag, New York, 2nd ed. (2003)
- [Kn06] Knörrer, H.: *Geometrie: Ein Lehrbuch für Mathematik- und Physikstudierende*. Vieweg, Wiesbaden, 2nd ed. (2006)
- [KP02] Krantz, S., Parks, H.: *A Primer of Real Analytic Functions*. Birkhäuser Advanced Texts: Basel Textbooks, Birkhäuser Boston Inc., Boston, 2nd ed. (2002)
- [Li73] Lieb, E. H.: *Convex trace functions and the Wigner-Yanase-Dyson conjecture*. Adv. in Math. 11, 267–288 (1973)
- [Ma07] Matúš, F.: *Optimality conditions for maximizers of the information divergence from an exponent family*. Kybernetika 43, 731–746 (2007)
- [Pe08] Petz, D.: *Quantum Information Theory and Quantum Statistics*. Theoretical and Mathematical Physics, Springer-Verlag, Berlin (2008)
- [Ro72] Rockafellar, R. T.: *Convex Analysis*. Princeton University Press, Princeton (1970)
- [We78] Wehrl, A.: *General properties of entropy*. Reviews of Modern Physics 50, 221–260 (1978)
- [We09] Weis, S.: *Exponential Families with Incompatible Statistics and Their Entropy Distance*. PhD Thesis, OPUS Erlangen-Nürnberg, [www.opus.ub.uni-erlangen.de/opus/volltexte/2010/1580](http://www.opus.ub.uni-erlangen.de/opus/volltexte/2010/1580) (2009)
- [Wi63] Wichmann, E. H.: *Density matrices arising from incomplete measurements*. J. Math. Phys. 4, 884–896 (1963)